



EUI WORKING PAPERS IN ECONOMICS

EUI Working Paper ECO No. 94/45

On the Differential Geometry of the Wald Test with Nonlinear Restrictions

FRANK CRITCHLEY,
PAUL MARRIOTT
and
MARK SALMON

P
30
UR

European University Institute, Florence

European University Library



3 0001 0016 5577 0



EUROPEAN UNIVERSITY INSTITUTE, FLORENCE

ECONOMICS DEPARTMENT

WP 330
EUR

EUI Working Paper ECO No. 94/45

**On the Differential Geometry of the Wald Test
with Nonlinear Restrictions ***

**FRANK CRITCHLEY,
PAUL MARRIOTT
and
MARK SALMON**

School of Mathematics and Statistics, University of Birmingham,
Department of Mathematical and Computing Sciences, University of Surrey and
Department of Economics, The European University Institute, Florence, Italy

This Version: November 17, 1994

* This paper is a completely rewritten and revised version of a paper originally circulated with the same title in 1989. We have received a number of comments on earlier versions of this paper but we would like to thank in particular Russell Davidson, Gene Savin and the anonymous referees. Simon Peters carried out the original monte carlo simulations.

BADIA FIESOLANA, SAN DOMENICO (FI)



All rights reserved.
No part of this paper may be reproduced in any form
without permission of the authors.

© Frank Critchley, Paul Marriott and Mark Salmon
Printed in Italy in December 1994
European University Institute
Badia Fiesolana
I – 50016 San Domenico (FI)
Italy

Abstract

In this paper we exploit the tools of Differential Geometry to analyse the finite sample lack of invariance of the Wald Statistic to algebraically equivalent reformulations of the null hypothesis. The Wald Statistic is shown, in general, to be an improper geometric quantity and hence is not invariant to reparameterisations of the statistical manifold in which it is being used. There is therefore little that can be done to rescue the Wald statistic from this sensitivity to the essentially arbitrary algebraic form in which the null hypothesis is expressed and the testing of nonlinear restrictions should be carried out using invariant approaches such as the Score or Likelihood Ratio procedures instead. The geometric approach also suggests an alternative invariant test based on the calculation of geodesic distances in curved manifolds. We show how this "Fisher Geodesic Statistic" may be easily calculated and applied in the case of testing nonlinear restrictions in the general linear model and also when it will coincide with the Wald Statistic. We are also able to extend the familiar inequalities relating the Wald, Score and Likelihood Ratio Statistics to the nonlinear case with the fundamental difference that the Fisher Geodesic Statistic takes the place previously occupied by the Wald statistic in the relevant inequality. The paper also provides an introduction to the methods of differential geometry and hopefully demonstrates its potential for econometricians.

1 Introduction

The Wald test, in different forms, is one of the most widely applied in Econometrics despite a fundamental deficiency in its finite sample behaviour when testing nonlinear restrictions. The problem manifests itself in a lack of invariance to algebraically equivalent reformulations of the null hypothesis. This sensitivity to the essentially arbitrary choice of the form of the restriction function means that any inference drawn is likely to be equally arbitrary given that the relevant finite sample distribution varies with the algebraic form chosen to express the null hypothesis¹. Gregory and Veall (1985) concluded their Monte Carlo study of a particular example by emphasising "the need for an analytical resolution to the problem of Wald test sensitivity". In this paper we aim to provide such an analysis by developing a clear geometric explanation of the invariance issue using the tools of differential geometry. In addition we derive from geometric arguments an invariant "Fisher Geodesic" statistic which is a natural, geometrically invariant analogue of the Wald Statistic and is easily applied in the case of testing nonlinear restrictions in the linear model.

There have been considerable developments at the interface between differential geometry and statistical inference recently; see for example, the review papers of Barndorff-Nielsen, Cox and Reid (1986) and Kass (1989) and the books by Amari (1985) or Murray and Rice (1993) and Barndorff-Nielsen and Cox (1994). An important motivating factor has been the power of geometric analysis which is particularly apparent when considering the issue of invariance. It is after all natural when designing inference procedures, to require that they should not depend upon the essentially arbitrary way in which we choose to label the density functions that constitute our models. In an exactly analogous way, geometry is concerned with those properties of spaces that do not depend upon a particular coordinate system or the parameterisation used to label its points. Thus both disciplines are concerned with quantities that are invariant under reparametrisation and hence it is natural that we should use the tools of differential geometry to analyse the lack of invariance of the Wald statistic.

Two, apparently independent, literatures relate to the results of this paper. The first has concentrated directly on the lack of invariance of the Wald statistic and references here include Gregory and Veall (1985), Lafontaine and White (1986), Breusch and Schmidt (1988), Phillips and Park (1988), Nelson and Savin (1988) and Dagenais and

¹Simply providing a size adjustment for the test does not remove the problem in this case as the power of such size adjusted tests will vary with the assumed form of the restriction function.

Dufour (1991) amongst others². One difficulty with interpreting this earlier body of work is that the non-invariant behaviour of the Wald Statistic has often only been numerically illustrated with particular examples using monte-carlo methods without providing a clear and general explanation or resolution of the problem.

The second body of work relevant to this paper relates to the question of testing (nonlinear) inequality restrictions; see for instance Gouriéroux, Holly and Monfort (1982), Gouriéroux and Monfort (1989), Kodde and Palm (1986) and Wolak (1989). Somewhat surprisingly the invariance problem, despite its relevance, has been almost completely overlooked in these papers although Gouriéroux and Monfort (1989) briefly mention it. Wolak(1989), for instance, derives the same statistic that we refer to as the Fisher Geodesic Statistic below but does so as an approximation to the Likelihood Ratio Statistic and relates it to the distance minimising criterion suggested by Kodde and Palm (1986). One benefit of the geometric approach we adopt below is that it clearly establishes the general principle that underlies these developments; that of using the geodesic distance between two distributions of interest as a basis for hypothesis testing. This justification is independent of any approximation to the Likelihood Ratio Statistic and thus represents what is apparently a distinct approach to hypothesis testing in general. It should however be made clear that we see a number of practical and theoretical difficulties in the application of the Geodesic principle in general nonlinear models and its relationship to the Likelihood Principle is not yet well understood. We are therefore only proposing at this stage, the use of the Geodesic Statistic in the case of testing nonlinear restrictions in the context of the general linear model. We note however that our geometric analysis of the failings of the Wald test is applicable beyond this relatively simple case and includes inference in situations that involve the more fundamental implications of nonlinearity³.

Phillips and Park (1988) provided the first general analysis of the Wald test invariance problem and showed, using Edgeworth expansions, how the structure of the Wald statistic varies with the form of the restriction function. Their approach explains, to a degree, the observed behaviour of the statistic for a given form of the null hypothesis as

²In addition papers by Nelson and Savin (1986), Griffiths, Hill and Pope (1987), and Calzolari and Fiorentini (1990) have discussed the directly related problem of estimating covariance matrices in nonlinear models using monte carlo methods. This latter issue of the effect of curvature on inference has of course a long history in statistics dating at least from the work of Beale (1960).

³Problems with the use of the Wald statistic in more general nonlinear or curved models, such as Tobit or Probit models, have been recognised for a number of years in the statistical literature, see in particular Hauck and Donner (1977, 1980), Moolgavkar and Venson (1987), Veath (1985) and more recently Le Cam (1990)).

the higher order terms in the expansions account for the finite sample deviations from the asymptotic distribution. They also provide correction factors that could in principle be used to indicate which particular parametric transformation of the restrictions describing the null will lead to a faster or better approximation of the finite sample distribution by the asymptotic distribution. However their analysis is limited to the $O(T^{-1})$ terms in the expansion and hence their corrections are similarly limited unlike the geometric analysis and Geodesic Test introduced in this paper and their correction factors vary with the true but unknown point on the null hypothesis. Following Phillips and Park, Dagenais and Dufour (1991) identified one essential part of the source of the lack of invariance of the Wald Statistic. However it is only when the geometric significance of their algebraic result is seen that a complete understanding of the issue becomes clear as explained below. Moreover the invariance (or lack thereof) of *alternative* test statistics, as shown in their monte-carlo simulations, can be straightforwardly rationalised from this broader geometric analysis. So our objective in this paper is to complement both these earlier studies by providing a geometric perspective that clarifies and extends the conceptual basis for their algebraic results.

The organisation and principal results of this paper are as follows. In Section 2 we introduce some basic differential geometry. Section 3 uses this geometry to explain the behaviour of the Wald statistic and other asymptotically equivalent statistics in broad generality. Since the Wald statistic is shown to be a hybrid geometric quantity we propose instead the use of a "Fisher Geodesic Statistic", FG, which emerges as a geometrically natural and hence invariant solution to the inference problem. We show that in the special case of linear models it is an explicit one-to-one function of the Likelihood Ratio Statistic, LR, and hence enjoys the same asymptotic properties under both the null and alternative. We also derive general inequalities linking the FG statistic to the LR and to the Score or Lagrange Multiplier, LM, statistics which extend the well known inequality relationships established by Berndt and Savin (1977) and Wolak(1988) for linear equality and inequality restrictions respectively. The geometry further indicates when the Wald and Fisher Geodesic statistics will coincide. Section 5 links the geometry to a graphical analysis which visually shows how the Wald test will behave for a particular choice of the algebraic form of the restriction function representing the null hypothesis. The Gregory and Veall (1985) example is used throughout and a Monte Carlo study is carried out which validates our geometric analysis. In Section 6 the considerably more complex issues raised by the use of geodesic statistics, in general, outside the context of the classical linear regression model are briefly discussed before we offer some conclusions.

2 The Wald Statistic

We consider a general statistical model $\mathcal{M} = \{p(x; \theta) | \theta \in \Theta\}$, as a set of probability density functions, indexed by θ given the observed data $x = (x_1, \dots, x_n)'$ on a random variable X . Let $\Theta \subseteq \mathbb{R}^p$ denote the parameter space of \mathcal{M} and $L(\theta; x) = \ln p(x; \theta)$ denote the corresponding log-likelihood function. We assume throughout that standard regularity conditions hold, see for instance Amari (1985, Section 2.1).⁴ The maximum likelihood estimator $\hat{\theta}$ of the unknown parameter θ is distributed, at least asymptotically, as multivariate normal, $N_p(\theta, I^{-1}(\theta))$ where $I(\theta)$ represents the Fisher Information Matrix given a sample of n , i.i.d. observations. We are concerned with testing the null hypothesis specified as the zero level set of a vector valued restriction function g . That is

$$H_0 = g^{-1}(0) \equiv \{\theta \in \Theta | g(\theta) = 0\}$$

where $g : \Theta \rightarrow \mathbb{R}^r$ ($1 \leq r \leq p$) is a vector of real-valued functions; we write $g(\theta) = (g_1(\theta), \dots, g_r(\theta))'$ and assume that Dg is of full rank throughout the domain.

The Wald statistic, $W(g)$, whose asymptotic distribution under the null is χ_r^2 , is then defined by:

$$W(g) = g(\hat{\theta})'[Dg(\hat{\theta})I(\hat{\theta})^{-1}Dg(\hat{\theta})']^{-1}g(\hat{\theta}).$$

As a quadratic form the Wald Statistic thus appears superficially to be a measure of distance, however as we show below it is *not* in general a geometric measure of distance between the null and the alternative hypotheses since it is not invariant to reparameterisation.

⁴In addition to the standard notation for derivatives we use two conventions that may be unfamiliar. The first indicates partial derivatives by $\partial_{i,j}$, for example, where the variables serving as arguments are identified by the context. The second convention is to omit the point of evaluation. Thus, we might use any of the expressions

$$(D^2g)_{i,j} = \partial_{i,j}g = \partial_{i,j}g(\theta) = \frac{\partial^2 g}{\partial \theta_i \partial \theta_j}(\theta)$$

when $g : \mathbb{R}^p \rightarrow \mathbb{R}$. Likewise any of

$$D_{\theta}g = Dg(\theta) = (\partial_j g_i) = \left(\frac{\partial g_i}{\partial \theta_j} \right)$$

when $g : \mathbb{R}^p \rightarrow \mathbb{R}^r$ has component functions $g_1(\theta), \dots, g_r(\theta)$.

3 The Geometric Measurement of Distance.

Dagenais and Dufour (1991) suggest three ways invariance may arise:

- (a) following a reparametrisation of the model,
- (b) or a reformulation of the null hypothesis,
- (c) or a smooth one-to-one transformation of the variables.

Transforming the variables implies a reparameterisation and provided no unknown parameters are involved in the transformation case (c) will in general imply case (a). In the context of the Wald test, reformulating the null hypothesis in effect chooses a reparametrisation, as we show below, and so (b) is also a special case of (a). Hence we shall concentrate throughout this paper on the need for a clear geometric explanation for a lack of invariance in the face of a reparametrisation of the model; case (a) above.

In this section we introduce sufficient differential geometry to be able to analyse the behaviour of the Wald test in the face of reparameterisation. The main objective is to explain a standard way used in differential geometry to measure the distance between points; the geodesic distance. It is important to recognise that the geodesic approach is applicable in both linear spaces, which covers standard linear statistical models *but also* in the nonlinear spaces in which many econometric inference problems fall and moreover it is by construction invariant to any reparametrisation of the inference problem. The Wald statistic will be shown in Section 4 to be a *non-invariant* approximation to this squared geodesic distance.

We define a space M to be a (p -dimensional) *manifold* if there exists an open subset of \mathbf{R}^p , U , and a map

$$\theta : M \rightarrow U \subset \mathbf{R}^p$$

such that θ is invertible and both θ and its inverse are smooth maps (where by smooth we mean infinitely differentiable, although in practice the existence of a finite number of derivatives will usually suffice) and we call θ a *parameterisation* of M .⁵ Our general statistical model $\mathcal{M} = \{p(x; \theta) | \theta \in \Theta\}$ represents an example of such a manifold and

⁵For a formal, and more general, definition of a manifold see Amari (1985, page 15).

hence we are concerned with measuring the distance between two statistical distributions in this space.

We shall throughout this section explain the essential geometric concepts using a simple but abstract example in \mathbf{R}^2 since it would substantially complicate the presentation to explain the same ideas in the function space defined by our statistical problem. Given that geometry is, in general, the study of those quantities which are invariant to reparametrisation, that is, it studies aspects of a manifold which are the same whether we work in the θ -parameterisation, or in some other parameterisation ξ , we start by considering the effects of reparametrisation.

Reparametrisation

Consider the subset of \mathbf{R}^2 which in Cartesian coordinates is given by

$$\mathbf{R}_+^2 = \{(\eta_1, \eta_2)' | \eta_1 > 0, \eta_2 > 0\}.$$

This same set can be equally well described in polar coordinates:

$$\mathbf{R}_+^2 = \{(r, \alpha)' | r > 0, 0 < \alpha < \pi/2\}.$$

and these two coordinate systems are related by the familiar equations:

$$r^2 = \eta_1^2 + \eta_2^2 \text{ and } \tan \alpha = \frac{\eta_2}{\eta_1}$$

and, conversely:

$$\eta_1 = r \cos \alpha \text{ and } \eta_2 = r \sin \alpha.$$

Since $(0, 0)' \notin \mathbf{R}_+^2$, we have that the change of coordinate functions from $(\eta_1, \eta_2)' \rightarrow (r(\eta_1, \eta_2), \alpha(\eta_1, \eta_2))'$ and back $(r, \alpha)' \rightarrow (\eta_1(r, \alpha), \eta_2(r, \alpha))'$ are mutually inverse and smooth.

Curves and tangent vectors.

Now using first the Cartesian, $(\eta_1, \eta_2)'$ -coordinate system, we can define a *parameterised path* or curve in \mathbf{R}_+^2 to be a smooth map

$$\gamma : [0, 1] \rightarrow \mathbf{R}_+^2$$

$$t \rightarrow \gamma(t) = (\eta_1(t), \eta_2(t))'$$

whose derivative is nowhere zero. Our geometric definition of distance will be based on the length of such curves. We can also define the same curve in polar coordinates where it will be given by

$$t \rightarrow \gamma(t) = (r(t), \alpha(t))'$$

where now

$$\eta_k(t) = r_k(r(t), \alpha(t))'$$

We will need the concept of a *tangent vector* to a curve in a manifold. In our example, the tangent vector to γ at the point $P = \gamma(t)$ is defined to be

$$\left(\frac{d\eta_1}{dt}(t), \frac{d\eta_2}{dt}(t) \right)'$$

for each $t \in [0, 1]$.

There are infinitely many curves which pass through any point of a manifold. However, in a given parameterisation of the manifold, there are certain curves which play an important role. In our example, let P be any point of \mathbf{R}_+^2 with coordinates $(P_1, P_2)'$. Then the curves

$$\gamma_1(t) = (P_1 + t, P_2)' \text{ and } \gamma_2(t) = (P_1, P_2 + t)'$$

pass through P and have tangent vectors at P given by $\partial_1 = (1, 0)'$ and $\partial_2 = (0, 1)'$ respectively. An important general result is that the set of all tangent vectors through a point P , in a p -dimensional manifold, is a p -dimensional real vector space referred to as the *tangent space* to M at P and denoted by TM_P . In our example, since $\{\partial_1, \partial_2\}$ are linearly independent, they form a basis for $T(\mathbf{R}_+^2)_P$. This is called the *natural basis* with respect to η -coordinates which generalises in an obvious way to give the natural basis $\{\partial_1, \dots, \partial_p\}$ for TM_P for any coordinate system θ .

Because of our requirement for invariance, it is important to know how the natural basis, and therefore how any tangent vector, transforms when we change from one

coordinate system θ to another, say ξ . In the example, if we change from Cartesian to polar coordinates, the tangent vector at P in $(r, \alpha)'$ coordinates is, by definition

$$\left(\frac{dr}{dt}(t), \frac{d\alpha}{dt}(t)\right)'$$

which is related to that in $(\eta_1, \eta_2)'$ -coordinates by the chain rule of differentiation.

Explicitly, we have

$$\frac{d\eta_k}{dt}(t) = \frac{d}{dt}\eta_k(r(t), \alpha(t)) = \frac{\partial\eta_k}{\partial r} \frac{dr(t)}{dt} + \frac{\partial\eta_k}{\partial\alpha} \frac{d\alpha(t)}{dt}$$

so that

$$\left(\frac{d\eta_1}{dt}, \frac{d\eta_2}{dt}\right)' = \begin{pmatrix} \cos \alpha(t) & -r(t) \sin \alpha(t) \\ \sin \alpha(t) & r(t) \cos \alpha(t) \end{pmatrix} \left(\frac{dr}{dt}, \frac{d\alpha}{dt}\right)'$$

and there is a similar inverse relation. In the general case, let $\{\partial_1, \dots, \partial_p\}$ and $\{\tilde{\partial}_1, \dots, \tilde{\partial}_p\}$ be the natural bases in the θ and ξ -parametrisations respectively. Then we find

$$\partial_i = \sum_{\alpha=1}^p B_{\alpha i} \tilde{\partial}_\alpha$$

and conversely

$$\tilde{\partial}_\alpha = \sum_{i=1}^p \bar{B}_{i\alpha} \partial_i$$

where the matrices with general elements $B_{i\alpha} = \frac{d\xi_\alpha}{d\theta_i}$ and $\bar{B}_{i\alpha} = \frac{d\theta_i}{d\xi_\alpha}$ are mutually inverse, thus in general $\frac{d\theta}{dt} = \bar{B} \frac{d\xi}{dt}$ and conversely $\frac{d\xi}{dt} = B \frac{d\theta}{dt}$. This combination of a definition in a particular coordinate system and a change of basis transformation rule is fundamental to the geometric approach.

Metric tensors.

In order to proceed with our geometric definition of distance, we have to be able to measure the length of a tangent vector, and to do so in a way that is invariant to reparametrisation. The length of a tangent vector in TM_P , and angles between tangent vectors in this space, can be determined by defining a symmetric, positive definite quadratic form on it. In a particular parameterisation θ , this form will have a matrix

representation G with respect to the natural basis $\{\partial_1, \dots, \partial_p\}$. Thus for each tangent space, TM_P , there will be a matrix $G = G(P)$ representing the quadratic form which is defined on that space.

The length invariance requirement will define a transformation rule between G and \tilde{G} , where the latter is the matrix representation of the same quadratic form with respect to the natural basis $\{\tilde{\partial}_1, \dots, \tilde{\partial}_p\}$ for the ξ -parameterisation. When we require also that $G(P)$ varies smoothly with the point P , we say that G is a *metric tensor* (for a general definition see Amari (1985, page 25)).

Using Cartesian coordinates (η_1, η_2) , we define the strictly positive quadrant of the plane to be *Euclidean* by taking G to be the 2×2 identity matrix for each $\eta \in \mathbf{R}_+^2$. Thus the length of the tangent vector to γ at $P = \gamma(t)$ is the nonnegative square root of:

$$\left\| \left(\frac{d\eta_1}{dt}, \frac{d\eta_2}{dt} \right) \right\|^2 = \left(\frac{d\eta_1}{dt}, \frac{d\eta_2}{dt} \right) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{d\eta_1}{dt} \\ \frac{d\eta_2}{dt} \end{pmatrix} = \left(\frac{d\eta_1}{dt} \right)^2 + \left(\frac{d\eta_2}{dt} \right)^2$$

Each tangent space then has the standard Euclidean norm.

Recall the relationship between the form of a tangent vector in Cartesian and polar coordinates,

$$\left(\frac{d\eta_1}{dt}, \frac{d\eta_2}{dt} \right)' = \bar{B} \left(\frac{dr}{dt}, \frac{d\alpha}{dt} \right)'$$

where

$$\bar{B} = \begin{pmatrix} \cos \alpha(t) & -r(t) \sin \alpha(t) \\ \sin \alpha(t) & r(t) \cos \alpha(t) \end{pmatrix}$$

The representation $\tilde{G}(r, \alpha)$ of the Euclidean metric in polar coordinates is then determined by the length invariance requirement that, for all tangent vectors:

$$\left(\frac{dr}{dt}, \frac{d\alpha}{dt} \right) \tilde{G} \begin{pmatrix} \frac{dr}{dt} \\ \frac{d\alpha}{dt} \end{pmatrix} = \left(\frac{d\eta_1}{dt} \right)^2 + \left(\frac{d\eta_2}{dt} \right)^2$$

Combining these two equations and given the form of \bar{B} we find that:

$$\tilde{G} = \bar{B}' \bar{B} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}$$

This generalises to the case of an arbitrary metric G at once to give the transformation rule

$$\bar{G} = \bar{B}' G \bar{B}$$

or conversely

$$G = B' \bar{G} B$$

where $B_{i\alpha} = \frac{\partial \xi_\alpha}{\partial \theta_i}$ and $\bar{B}_{\alpha i} = \frac{\partial \theta_i}{\partial \xi_\alpha}$ as before.

There is, however, a very important difference between $G(\eta_1, \eta_2)$ and $\bar{G}(r, \alpha)$. The former is a *constant* for all tangent spaces while the second is a *function of r* and so varies with position. This essentially arises because r and α are not just affine functions of η_1 and η_2 . The impact of this difference between a constant metric and a non constant metric will be felt in our analysis of the Wald test below. Notice that the norm of the tangent vector defined by

$$\left\| \frac{d\gamma(t)}{dt} \right\|^2 = \frac{d\gamma(t)'}{dt} G \frac{d\gamma(t)}{dt}$$

is invariant by construction. Note that the matrix form of the Euclidean metric is diagonal in both coordinate systems considered. This reflects the fact that, not only are the $\eta_1 = \text{constant}$ curves everywhere orthogonal to the $\eta_2 = \text{constant}$ curves, but also so are the constant curves of r and α . We are now in a position to define the Geodesic Distance between two points in M .

Geodesic distances.

Let $\gamma(t) : [0, 1] \rightarrow M$ be a path in M . Using a metric tensor G on M we have an invariant definition of the length of each tangent vector $\frac{d\gamma}{dt}(t)$, $t \in [0, 1]$. Regarding t as denoting time, we can think of this length as the speed with which the point $\gamma(t)$ moves across M . It would then be natural to define the length of the curve γ as being the distance travelled by $\gamma(t)$ as t moves from 0 to 1. In other words to define the distance travelled to be the integral of the speed over the time taken.

Formally, the length of the curve γ from $\gamma(0)$ to $\gamma(1)$ is defined to be the integral

$$\int_0^1 \left\| \frac{d\gamma(t)}{dt} \right\| dt. \quad (1)$$

In fact, more generally, we allow an integration over any interval where γ is defined. As the integrand is by construction invariant to any reparametrisation, so is the value of the integral itself and thus we have an invariant definition of path length.

For any two points P and Q on the manifold M , there are infinitely many paths joining them. Intuitively, we may then define the *geodesic distance* from P to Q to be the minimum such path length, and a path which attains this minimum to be a *geodesic* (curve or path) joining P and Q . Since we have just seen that path length is invariant to reparametrisation, so are geodesic distances and geodesic curves. For the applications in this paper, this definition will suffice but notice that for more complex problems this definition does not address several important issues such as existence and uniqueness. The standard geometric definition is made via a local path length minimisation, rather than the global one used above; for details, see Dodson and Poston (1977).

As an example, consider the points $P = (1, 1)'$ and $Q = (2, 2)'$ in \mathbf{R}_+^2 in the η -parameterisation. To find the geodesic curve joining P and Q , and its length, when the Euclidean metric is used, we must solve the following problem. Find $\gamma(t) = (\eta_1(t), \eta_2(t))'$ which minimises

$$\int_0^1 \left\{ \left(\frac{d\eta_1(t)}{dt} \right)^2 + \left(\frac{d\eta_2(t)}{dt} \right)^2 \right\}^{\frac{1}{2}} dt$$

subject to the constraints that $\eta_1(0) = \eta_2(0) = 1$ and $\eta_1(1) = \eta_2(1) = 2$.

This is a classical problem in the calculus of variations with the solution:

$$\eta_1(t) = 1 + t, \quad \eta_2(t) = 1 + t.$$

Thus the geodesic curve joining P and Q is affine in the η -coordinates and the geodesic distance between the two points is given by

$$\int_0^1 \sqrt{2} dt = \sqrt{2}.$$

In a manifold, M , with a coordinate system θ for which the metric is constant for all tangent spaces the above result generalises in the following way. The geodesic curve between $\theta = (\theta_1, \dots, \theta_p)'$ and $\theta^* = (\theta_1^*, \dots, \theta_p^*)'$ is given by

$$\gamma_i(t) = (1-t)\theta_i + t\theta_i^* \quad 0 \leq t \leq 1$$

as an *affine* combination of the endpoints. Further, there is an explicit closed form solution for the geodesic distance. If the metric tensor is given by the constant matrix G then the geodesic distance is given by

$$\int_0^1 ((\theta^* - \theta)'G(\theta^* - \theta))^{\frac{1}{2}} dt = ((\theta^* - \theta)'G(\theta^* - \theta))^{\frac{1}{2}} \quad (2)$$

the square root of a quadratic function of the difference in the endpoints. Although the metric is defined as a quadratic form on a tangent space it does in *this very special case* of a constant metric, G , also have a direct interpretation on the manifold.

However if we move away from the constant metric case, induced for example by a nonlinear reparametrisation, these results will not hold. Thus in general geodesic curves are not simply affine combinations of their endpoints. Moreover, squared geodesic distances are not in general a quadratic function of differences in endpoints determined by a single quadratic form. It is therefore important to notice for our analysis below that such quadratic functions are not in general invariant to reparametrisation. These difficulties lie at the heart of the problem with the Wald test.

We can illustrate these remarks by again considering polar coordinates in \mathbf{R}_+^2 and repeating the above exercise. In $(r, \alpha)'$ -coordinates $P = (\sqrt{2}, \pi/4)'$ and $Q = (2, \pi/4)'$. By invariance, we know that the geodesic curve joining them is, in $(r, \alpha)'$ -coordinates, given by $r^2(t) = 2(1+t)^2$ and $\alpha(t) = \pi/4$. This is not an affine function of the endpoints. Using invariance again, the squared geodesic distance from one general point $P = (r_1, \alpha_1)'$ to $Q = (r_2, \alpha_2)'$, will be

$$\{\eta_1(Q) - \eta_1(P)\}^2 + \{\eta_2(Q) - \eta_2(P)\}^2 = r_1^2 + r_2^2 - 2r_1r_2 \cos(\alpha_1 - \alpha_2)$$

which is *not* a quadratic form in $(r_2 - r_1, \alpha_2 - \alpha_1)'$. In particular, recalling the matrix representation of the Euclidean metric in polar coordinates, it is not of the form

$$(r_2 - r_1, \alpha_2 - \alpha_1) \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix} \begin{pmatrix} r_2 - r_1 \\ \alpha_2 - \alpha_1 \end{pmatrix}$$

for any r . Thus a fixed quadratic form using the metric at a single point does not give an invariant measure of distance in the manifold.

We will need to consider an important generalisation of the definition of the distance between two points; that is the distance between a point P and a line, or more generally, a submanifold, N . We could, for instance, regard the point as that representing the unrestricted MLE and the submanifold N as representing the null hypothesis. We can define this distance to be the minimum of those between P and Q , where $Q \in N$. If Q_0 is the point at which this minimum is attained then the geodesic joining Q_0 and P cuts N orthogonally, see Spivak (1981).

Finally given the discussion above we can see that the existence of coordinates which provide a metric with a constant representation simplifies geometric calculations considerably. Such coordinates are called *affine* and when they exist they are unique up to a non-singular affine reparametrisation. There is an important characterisation of this case. Given (M, G) , an arbitrary manifold M with a metric tensor G , we can define a tensor R called the Riemann-Christoffel curvature tensor (see Amari (1985, page 46)). This is a function of G and its derivatives with the property that $R(\theta) \equiv 0$ if and only if a set of affine coordinates exists for G . Being a tensor, R vanishes in one coordinate system if and only if it does in all coordinate systems.

Non-affine coordinates may arise in two ways. First, through a nonlinear reparameterisation of a set of affine coordinates, as in the polar coordinate example above, or because R does not vanish at some point and so all coordinate systems are necessarily non-affine. In Section 4 below, we are largely (although not exclusively) concerned with normal linear model analysis which falls into the first category. The general case is discussed in Section 6.

4 Wald test Geometry and the Fisher Geodesic Statistic.

Recalling the form of the Wald statistic for a restriction function g given in Section 2 by

$$W(g) = g(\hat{\theta})' [Dg(\hat{\theta}) I(\hat{\theta})^{-1} Dg(\hat{\theta})']^{-1} g(\hat{\theta}),$$

we now interpret the statistic in geometric terms and use the basic differential geometry given above to explain the lack of invariance to a reformulation of the null hypothesis.

The first geometric observation, going back at least to Rao (1945), is that the Fisher information matrix is a metric tensor. It is positive definite and symmetric. Let F and I be the matrix forms of the Fisher information at a given point in ξ and θ -coordinates respectively. Then, when expectations are taken with respect to the distribution for the

random variable X at that point, we have $1 \leq a, b \leq p$:

$$F_{ab} = E\left(\frac{\partial l}{\partial \xi_a} \frac{\partial l}{\partial \xi_b}\right) = E\left(\sum_{i=1}^p \sum_{j=1}^p \frac{\partial \theta_i}{\partial \xi_a} \frac{\partial l}{\partial \theta_i} \frac{\partial \theta_j}{\partial \xi_b} \frac{\partial l}{\partial \theta_j}\right) = \sum_{i=1}^p \sum_{j=1}^p \bar{B}_{ia} \bar{B}_{jb} I_{ij}$$

as required. Thus the Fisher information can be used to measure the lengths of vectors in the tangent space where it is evaluated.

The second observation is that if the Fisher information matrix is constant in one parameterisation then it need not necessarily be so in some other nonlinear reparameterisation. This corresponds to the example in the previous section where the Euclidean metric is constant (in fact the identity matrix) in Cartesian coordinates but varies in polar coordinates. We note that in the normal linear model $y = X\beta + \epsilon$, the information matrix $I(\beta) = \sigma^{-2}(X'X)$, is constant in β but in any coordinates which are nonlinear functions of β it will be nonconstant.

Our third observation is that the Wald statistic can be seen to involve precisely this type of nonlinear reparameterisation. One well recognised advantage of using the Wald statistic as opposed either the LM or LR statistics in the case of nonlinear restrictions is that it is not necessary to estimate the restricted model since only the unrestricted maximum likelihood estimate is used. Considering the form of the Wald test above we can see it uses, with this nonlinear null hypothesis, the restriction function g itself effectively as a set of parameters to make the contrast between the restricted and unrestricted cases in place of directly using the relevant parameter estimates. It compares the unrestricted estimated value of the " g " function with the " g -value" for any point on the null, taken to be zero. It is important to realise that the value of g will generally not precisely fix the position of the unrestricted and restricted parameter estimates as the g -value provides only a *partial parameterisation* of the parameter space unless the number of restrictions equals the dimension of the parameter space. In the case of linear restrictions in the normal linear model there is no effective loss of information given the marginalisation properties of normal distributions but in the case of nonlinear restrictions this partial reparameterisation will not be sufficient. In order to analyse this case it will be useful to extend the partial parameterisation induced by g to a full parameterisation as we can then directly see how a change in the form of the restriction function corresponds to a nonlinear reparameterisation.

The partial parameterisation g can be extended to a full parameterisation by introducing new coordinates k where k completes the coordinate system; $\theta \rightarrow \xi = (g, k)'$,

where $k(\theta) \in \mathbb{R}^{p-r}$. This full parameterisation will exist in a neighbourhood of the null hypothesis, technically in a *tubular neighbourhood*, see Spivak (1981). This result follows directly from the Implicit Function Theorem and uses the full rank of g .

Our fourth observation is that the Wald statistic ignores all information contained in the k coordinates.

The Wald statistic involves the inverse of the Fisher information matrix at $\hat{\theta}$ the unrestricted maximum likelihood estimate. Inverting the matrix form of the equation linking I and its reparameterised form F we find, writing $\hat{\xi} = \xi(\hat{\theta})$ and $\hat{F} = F(\hat{\xi})$, that:

$$\hat{F}^{-1} = \begin{pmatrix} Dg(\hat{\theta})I(\hat{\theta})^{-1}Dg(\hat{\theta})' & Dg(\hat{\theta})I(\hat{\theta})^{-1}Dk(\hat{\theta})' \\ Dk(\hat{\theta})I(\hat{\theta})^{-1}Dg(\hat{\theta})' & Dk(\hat{\theta})I(\hat{\theta})^{-1}Dk(\hat{\theta})' \end{pmatrix}$$

where we have used the relation

$$\bar{B}^{-1} = B = \begin{pmatrix} Dg \\ Dk \end{pmatrix}$$

This formula holds for an arbitrary extension of g to (g, k) -coordinates. It is possible to choose $k(\theta)$ so that, at $\hat{\theta}$, the k -constant lines are always orthogonal to the g -constant lines and hence the matrix form for \hat{F}^{-1} is block diagonal.⁶

$$\hat{F}^{-1} = \begin{pmatrix} Dg(\hat{\theta})I(\hat{\theta})^{-1}Dg(\hat{\theta})' & 0 \\ 0 & Dk(\hat{\theta})I(\hat{\theta})^{-1}Dk(\hat{\theta})' \end{pmatrix}$$

Clearly the Wald statistic, $W(g)$, is a quadratic form that only involves the top left-hand corner of the above partitioned matrix \hat{F} . Writing $\hat{\xi}' = (g(\hat{\theta}), k(\hat{\theta}))'$ and $\hat{\xi}_0' = (0, k(\hat{\theta}))'$, we see that

$$W(g) = (\hat{\xi} - \hat{\xi}_0)' \hat{F} (\hat{\xi} - \hat{\xi}_0), \quad (3)$$

which is the same form as the constant metric case, (2), of the general Geodesic distance given in (1). The metric \hat{F} is fixed given the particular value for $\hat{\theta}$.

This brings us to the heart of our diagnosis. Our last observation is that, viewed geometrically, the Wald statistic is a hybrid quantity, having features of two differing geometric quantities:

⁶Since the Wald statistic only involves this one point there is no loss of generality here.

- (i) The squared length of a vector in a tangent space,
and:
- (ii) a squared distance between two points in a manifold.

Individually both of these are genuine geometric objects in that they are completely invariant to reparametrisation. Unfortunately, the Wald statistic does not in general coincide with either of these two forms. It is not of the form (i), as $(\hat{\xi} - \hat{\xi}_o) = (g(\hat{\theta}) - g(\theta_o), 0)$ does not transform as a tangent vector and so $(\hat{\xi} - \hat{\xi}_o)$ does not lie in the tangent space. It is not a quantity of the second form as it uses a fixed metric \hat{F} where to be a squared distance measure in a manifold F would need in general to vary with ξ for the reasons explained in Section 3. The two cases only coincide for the special case of a constant metric and not in general. As a consequence, the Wald statistic is not a genuine geometric object and hence does not remain invariant when a reformulation of the null induces a nonlinear reparametrisation.

The geometric point of view immediately suggests a resolution in that the obvious thing to do is replace the hybrid quantity $W(g)$ by one of the pure geometric forms (i) or (ii); this will at least automatically guarantee invariance to reparametrisation. If we use approach (i) we find the score statistic or the $C(\alpha)$ statistic recently examined in this context by Dagenais and Dufour (1991). If we use approach (ii), we arrive at the squared geodesic distance between $\hat{\theta}$ and the nearest point on the null hypothesis H_0 and, if we use the Fisher information matrix as the metric tensor, it is not unreasonable to call this a Fisher Geodesic statistic and denote it by $FG(\hat{\theta}, H_0)$ or, simply, FG . Notice that the Likelihood Ratio statistic is invariant because it is simply a comparison of the values of the likelihood at two points in the manifold.

In what follows we note some of the properties of this Fisher Geodesic statistic in the normal linear model and then in Section 5 apply it to the Gregory and Veall (1985) example, both as a statistic in its own right and as a geometrically natural reference point for the Wald statistic. In Section 6 we briefly consider the general use of geodesic statistics outside the context of the general linear model.

Recall that in the classical linear regression model $y = X\beta + \epsilon$ we have $I(\beta) = \sigma^{-2}(X'X)$ and because the information matrix does not depend upon β , the geodesic curves are just straight lines in β -space (in the usual Euclidean sense) and the Fisher Geodesic statistic collapses to

$$FG = \min_{\beta \in H_0} \{(\beta - \hat{\beta})' \sigma^{-2} (X'X)(\beta - \hat{\beta})\}.$$

When σ^2 is unknown we replace it by the unrestricted maximum likelihood estimate $\hat{\sigma}^2$ to obtain the operational statistic \hat{FG} . If RSS denotes the residual sum of squares under the restriction $\beta \in H_0$ and if LR denotes the likelihood ratio statistic for this hypothesis, then the simple identity $RSS = \hat{\sigma}^2(n + \hat{FG})$ establishes that

$$LR = n \ln(1 + n^{-1} \hat{FG})$$

Thus, in this case, the geodesic statistic is an explicit one-to-one function of the likelihood ratio statistic.

Moreover, since $n \ln(1 + n^{-1}x) \sim x$ as $n \rightarrow \infty$, we see that LR and \hat{FG} will agree asymptotically under both the null and alternative. In particular they will have the same known asymptotic χ^2 distribution. Now, if we let

$$LM = \frac{n\hat{\sigma}^2 \hat{FG}}{RSS}$$

denote the Lagrange Multiplier statistic for H_0 then using $e^x \geq (1+x)$, we see that the following inequalities hold for any null hypothesis H_0 , which in particular need not be linear in β :

$$LM \leq LR \leq FG \tag{4}$$

These inequalities may be compared with those established between the Wald, LR and LM statistics for testing linear restrictions and inequality constraints by Berndt and Savin (1977), Evans and Savin (1982) and Wolak (1988). Hence we see that, although equivalent asymptotically, in finite samples the Fisher Geodesic test will, in normal linear models, reject H_0 whenever the Likelihood Ratio test does if both statistics are referred to the same distribution.

Finally, in the following proposition, we provide necessary and sufficient conditions under which the Fisher Geodesic statistic coincides with the Wald statistic. We note that these conditions are satisfied for *linear* formulations of linear restrictions on regression parameters in the normal linear model and thus, in a sense, the Fisher Geodesic test can be seen as a direct generalisation of the Wald test. Notice however that Wald

tests based on a *nonlinear* formulation of linear restrictions do not obey the conditions. We have the following result:

Proposition 4.1 Given the basic model of Section 2, under standard regularity conditions, the Wald statistic based on g coincides with the Fisher geodesic statistic if:

- (i) the Fisher information matrix does not depend upon θ . i.e.

$$\forall \theta \in \Theta, I(\theta) = I_0$$

for some symmetric, positive definite matrix I_0 .

and

- (ii) H_0 is the zero set of $g : \Theta \rightarrow \mathbf{R}^r$ defined by $g(\theta) = A\theta + b$ with A having full row rank r .

Proof. We find FG by essentially geometric considerations. Let R be the unique symmetric positive definite square root of I_0 . Then, under the reparametrisation $\theta \rightarrow \eta = R\theta$, the Fisher information matrix becomes the identity, that is to say η -space is Euclidean. Let $\hat{\eta}$ denote $R\hat{\theta}$. Since geodesic distances are invariant to reparametrisation, we have that $FG = (\hat{\eta} - \hat{\eta}_0)'(\hat{\eta} - \hat{\eta}_0)$ in which $\hat{\eta}_0$ is the (Euclidean) orthogonal projection of $\hat{\eta}$ onto $H_0 = \{\eta \in \mathbf{R}^p | C\eta + b = 0\}$ where $C = AR^{-1}$. Now $(\hat{\eta} - \hat{\eta}_0)$ is orthogonal to $\{(\eta - \hat{\eta}_0) | \eta \in H_0\}$. But as $\hat{\eta}_0 \in H_0$, this set is just the null space of C . In other words, $(\hat{\eta} - \hat{\eta}_0)$ lies in the range space of C' . Writing $(\hat{\eta} - \hat{\eta}_0) = C'\lambda$ and imposing the condition that $\hat{\eta}_0 \in H_0$, we find that $\lambda = (CC')^{-1}(C\hat{\eta} + b)$. Hence, as required

$$FG = g(\hat{\theta})'[AI_0^{-1}A']^{-1}g(\hat{\theta}).$$

5 An application: The Gregory and Veall example.

We now demonstrate how the geometric approach can be applied to examine the behaviour of the Wald statistic and the Fisher Geodesic statistic using the example due to Gregory and Veall (1985). They considered a normal linear model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i \quad (i = 1, \dots, n)$$

with $\epsilon_i \sim N(0, \sigma^2)$ and found large differences between the performance of two Wald statistics, $W(g^A)$ and $W(g^B)$, based on the following alternative formulations of the same null hypothesis:

$$H_0^A : g^A = 0 \quad \text{where} \quad g^A(\beta_1, \beta_2) = \beta_1 - \beta_2^{-1}$$

and

$$H_0^B : g^B = 0 \quad \text{where} \quad g^B(\beta_1, \beta_2) = \beta_1 \beta_2 - 1$$

As they are the only parameters involved in H_0 , attention focuses on the subvector $(\beta_1, \beta_2)'$ which we denote by β and assume to be strictly positive.

The monte carlo experiments presented below are based on $\sigma^2 = 1$ and a similar regression model design to the one used by Gregory and Veall except that we decided to use a constant $n \times 2$ matrix $X = (x_{ij})$ across replications. This corresponds to one realisation of the stochastic process used by Gregory and Veall but provides a constant framework for comparison across replications. Of course, X still varies with sample size, however the regressors were scaled in order to achieve similar design R^2 's across different series lengths. Although not essential to our geometric analysis, for clarity of exposition we also rotated the design so that the columns of X are orthogonal giving

$$I(\beta) = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}.$$

Since it is convenient to work in Euclidean coordinates we define $\eta_1 = \sqrt{a}\beta_1$ and $\eta_2 = \sqrt{b}\beta_2$, so that η -space is Euclidean.

5.1 A Monte Carlo Study.

The normal linear model described above was simulated with 10,000 replications for each case with three sample sizes of 20, 50 and 100 observations. The (a, b) values in the information matrix, $I(\beta)$, were taken respectively to be (24.6, 111.5), (285.2, 26.5) and (578.4, 124.7). The two versions of the Wald statistic corresponding to g^A and g^B were computed together with the Lagrange Multiplier, Likelihood Ratio and Fisher Geodesic statistics. These, asymptotically $\chi^2(1)$ statistics, were then compared at the 1% and 5% levels at five points on the null hypothesis such that the product of β_1 and β_2 is unity.

$$(\beta_1, \beta_2) = \begin{cases} (10, 0.1) \\ (2, 0.5) \\ (1, 1) \\ (0.5, 2) \\ (0.1, 10) \end{cases}$$

Table 1 indicates the number of times H_0 was rejected in each case (H_0 true). A standard binomial calculation shows that the monte carlo standard errors are ± 10 and ± 22 at the 1% and 5% levels respectively.

Table 1
Gregory and Veal Simulations
Rejection frequencies on
the Null given 10,000 replications

β_1	β_2		$n = 20$		50		100	
			1%	5%	1%	5%	1%	5%
10	0.1	R^2	0.9923		0.9983		0.9983	
		WA	1448	1960	2666	3288	1292	1829
		WB	272	836	165	655	144	568
		LM	97	567	107	556	110	526
		LR	177	709	139	610	121	548
		FG	274	838	165	656	141	568
2.0	0.5	R^2	0.8692		0.9590		0.9595	
		WA	276	664	612	1026	310	706
		WB	295	870	166	660	130	571
		LM	96	557	104	561	107	531
		LR	189	714	131	617	116	551
		FG	276	850	168	668	131	575
1.0	1.0	R^2	0.8775		0.8642		0.8766	
		WA	280	848	289	658	147	550
		WB	313	868	159	660	135	594
		LM	105	576	100	550	100	548
		LR	215	726	129	608	110	564
		FG	288	854	168	645	127	589
0.5	2.0	R^2	0.9597		0.7836		0.8667	
		WA	330	893	146	637	135	565
		WB	325	891	202	728	139	565
		LM	111	613	113	565	109	526
		LR	201	747	144	611	121	539
		FG	330	893	180	680	131	562
0.1	10.0	R^2	0.9983		0.9819		0.9221	
		WA	326	889	185	706	138	542
		WB	326	888	184	703	136	540
		LM	118	606	104	604	109	493
		LR	210	747	147	655	126	516
		FG	326	889	185	706	138	542

Focussing on the larger sample sizes where the asymptotic results may be expected to apply, we note the following features. Under H_0 ,

- (1) The Fisher Geodesic and Likelihood Ratio tests perform comparably throughout.
- (2) The $W(g^A)$ test, WA, performs well for large β_2 , but increasingly badly as β_2 de-

creases. At $\beta_2 = 0.1$, its behaviour is extremely poor. When $\beta_2=10$ its performance is effectively the same as the $W(g^B)$ and Fisher Geodesic tests. The other test statistics seem relatively unaffected by variations in β_2 .

- (3) The $W(g^B)$ test performs relatively well and closely follows the Fisher Geodesic test.
- (4) We note the systematic inequality between the Fisher Geodesic, the Likelihood Ratio and the Lagrange Multiplier statistic derived in the previous section.

To pursue the geometric analysis of these monte carlo results, we introduce a convenient choice of k in the parameterisation $\theta \rightarrow \xi = (g', k')'$ as discussed in Section 4 in which the form of the Fisher information matrix is diagonal. Such a parameterisation will be called *Fisher orthogonal*. The proofs of the following propositions are given in the appendix.

Proposition 5.1. For any smooth real valued restriction function g there exists a function $k : \mathbf{R}^2 \rightarrow \mathbf{R}$ such that in (g, k) -coordinates the Fisher information matrix is of the form

$$F(\xi) = \begin{pmatrix} f_{11} & 0 \\ 0 & f_{22} \end{pmatrix}$$

in some neighbourhood of the null hypothesis. Furthermore any two such functions, k and k^* will have the same level sets.

Taking the Fisher geodesic statistic as our benchmark for evaluating the performance of the Wald statistic we expect, following our analysis in Section 4, that the Wald test will differ from the Fisher Geodesic test, if $F(\xi)$ varies substantially near $\hat{\theta}$ and we can develop measures of this variation where g and k are scalars as in the Gregory and Veall example. It would be natural to analyse the change in the form of the information matrix in terms of the functions f_{11} and f_{22} . However since there is some degree of choice in the function k we need to ensure that the diagnostics used are invariant to different choices of k .

Proposition 5.2. For any g, g_1, g_2 ; $f_{11}(g, k)$ and $\frac{f_{22}(g_2, k)}{f_{22}(g_1, k)}$ are invariant to different choices of k .

Further the quantity

$$m(g, k) = \frac{\partial}{\partial g} \ln f_{22}(g, k)$$

is an appropriate measure of the variation in f_{22} and is invariant to different choices of the Fisher orthogonal coordinates k .

Thus we have two diagnostics that can be used to indicate when the Wald test may behave badly i.e. when:

- (i) when there is large variation in f_{11} near $\hat{\beta}$,
and/or
- (ii) when large values of $|m(g, k)|$ are found near $\hat{\beta}$.

We next calculate these two quantities for g^A and g^B , and use them to predict the performance of $W(g^A)$ and $W(g^B)$ as actually observed in the monte-carlo experiment. We need to calculate a set of Fisher orthogonal coordinates for the Gregory and Veal example for both the g^A and g^B forms of the restriction function and the form of the metric in these coordinate systems.

Proposition 5.3. For the g^A restriction function we may take

$$k^A(\beta_1, \beta_2) = d^2\beta_1 - \frac{1}{3}\beta_2^3$$

where $d = \sqrt{\frac{a}{b}}$, so that (g^A, k^A) is a Fisher orthogonal coordinate system. Similarly for the g^B form of restriction function we may take

$$k^B(\beta_1, \beta_2) = \frac{1}{2}(\beta_1^2 - d^{-2}\beta_2^2).$$

Proposition 5.4. For g^A :

$$f_{11}^A = a^{-1} + b^{-1}\beta_2^{-4}, \quad f_{22}^A = b(d^2 + \beta_2^4)^{-1}$$

and

$$m^A(g, k) = -\frac{4d^2\beta_2^5}{(d^2 + \beta_2^4)^2}$$

and for g^B we have :

$$f_{11}^B = a(\beta_2^2 + d^2\beta_1^2)^{-1}, \quad f_{22}^B = \frac{4ad}{d + 4\beta_2^2}$$

and

$$m^B(g, k) = -\frac{4d^2\beta_1\beta_2}{(\beta_2^2 + d^2\beta_1^2)^2}$$

The above measures indicate the variation of the metric tensor in a region of β . In this two dimensional example the variation can also be seen graphically. In Figures 1 and 2 we plot the (g, k) -coordinates, for g^A and g^B respectively, in a neighbourhood of the null hypothesis. In each figure g -constant lines are plotted on both sides of the null in the (η_1, η_2) Euclidean coordinates introduced above (on page 19). In these new coordinates the orthogonality of the (g, k) -coordinate system has the straightforward visual interpretation that the g -constant lines will cut the k -constant lines orthogonally in each diagram.

Either the diagnostics of Proposition 5.4 or Figures 1 and 2 can be used to explain the monte carlo results. Firstly considering the g^A restriction function; for large values of β_2 , on the null, the metric in (g, k) -coordinates will remain fairly constant since $f_{11}^A \rightarrow a^{-1}$, a constant, and $m^A \rightarrow 0$ as $\beta_2 \rightarrow \infty$. This would indicate a close agreement between the Geodesic and Wald statistics for these values of β_2 . Graphically the region where β_2 is large on the null corresponds to region (1) in Figure 1. Here the (g, k) -coordinates are visually indistinguishable from Euclidean. The results from the monte carlo experiment exactly agree with these predictions.

For values on the null where β_2 is small we see large variation in the metric. There is a singularity in f_{11}^A as $\beta_2 \rightarrow 0$. Thus we would expect a large difference between the Wald and the Geodesic statistics in this case. In the figure this singularity in the metric lies in region (2) where the singularity in the coordinate system is obvious. This again exactly corresponds to the behaviour observed in the monte carlo experiment.

For the restriction function g^B there is much less variation in the diagnostic functions as β_1 and β_2 vary. First we see considerable symmetry in the diagnostics as $\beta_2 \rightarrow 0$ or ∞ . In both cases $m^B \rightarrow 0$ and also f_{11}^B does not have a singularity. Visual inspection of Figure 2 shows that the (g, k) -coordinate system is always approximately Euclidean, especially in the tails of the hyperbola, $\beta_1\beta_2 - 1 = 0$ which is the null hypothesis. These observations imply that there should always be a reasonable agreement between the Geodesic and the Wald statistics in this case, particularly for large or small

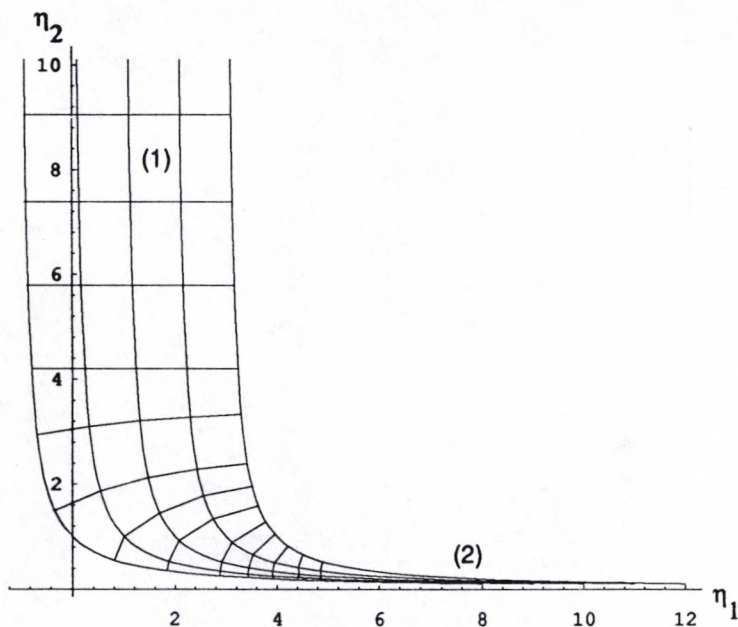


Figure 1: The g^A and k^A coordinate system

values of β_2 . This is again in exact agreement with the monte carlo results.

6 Geodesic distances in Statistics.

In the case of the linear model considered above we used a definition of geodesic distance which was induced by the Fisher information metric. This defined an invariant distance function on a parametric family which can be used as a test statistic. In this section we briefly discuss the general properties of this type of test statistic and other related statistical distance measures.

The use of geodesics in statistical inference has recently been recognised, in particular in the work of Amari (1985) and Barndorff-Nielsen (1989). This literature uses the idea of a geodesic which is defined by an *affine connection* rather than a metric tensor. This more general, though less intuitive, definition of a geodesic is important in the

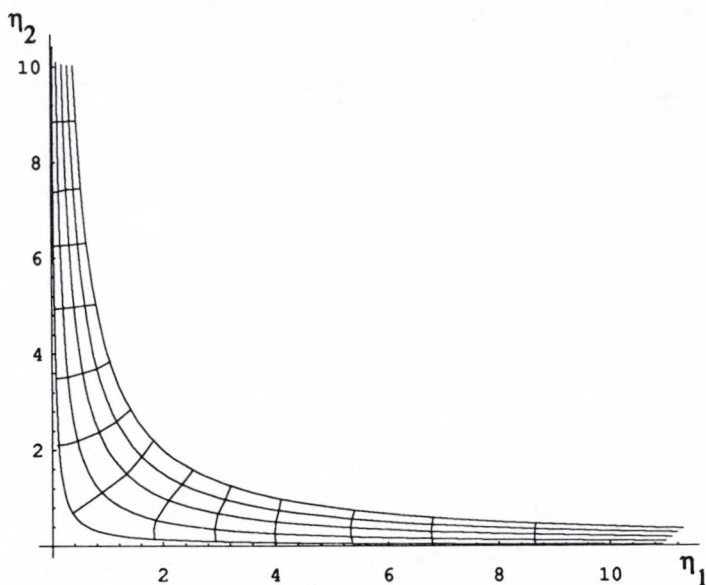


Figure 2: The g^B and k^B coordinate system

general theory of differential geometry and in particular in its applications to Physics. For a good reference to these more general ideas see Dodson and Poston (1977). For the present we concentrate on metric tensor based geodesics since they have the added structure that there is always an associated *geodesic distance*. A geodesic defined by an affine connection does not in general have either the length minimising property or the corresponding distance measure and it is intuitively reasonable to expect that these distances might prove to be useful test statistics. It is as yet an open question if such a geometrically based test statistic has good statistical properties in general but certain important relationships with well-known statistical objects do exist.

In Critchley, Marriott and Salmon (1993) it is shown how the geometry of Amari, which is defined using affine connections, can be viewed as a metric tensor based geometry. To do this though it is necessary to define the concept of a *preferred point metric*, $g^\phi(\theta)$. The *preferred point*, ϕ , is some point in the manifold which is singled out as special in the geometry. Statistically it may be thought of as the data generation process or some estimate of it, and for each potential choice of ϕ , $g^\phi(\theta)$ is a metric tensor. There

are a number of statistically natural choices for a preferred point metric structure on a manifold, each containing some particular statistical information. It is shown in Critchley, Marriott and Salmon (1994a) that under a condition called *total flatness* these different choices reduce to one, which agrees with the Fisher information metric used above. The linear model discussed above is one of the few families which is totally flat. This gives an important theoretical justification for the use of the Fisher metric in our analysis.

For more general models a distinction has to be drawn between the different possible preferred point metrics. In the case of full exponential families the different choices can be seen to correspond to the set of α -geometries defined by Amari and δ -parameterisations defined by Kass (1984) and Hougaard (1982). For a discussion of this issue see Critchley, Marriott and Salmon (1994b).

There already exist distance measures in statistics. Chentsov (1972) and Amari (1985) consider a class of distance measures called *divergences*. Included in this class are the Kullback-Leibler and Hellinger distances. These measures do not possess the symmetry condition needed in the strict mathematical sense of distance, however they are natural for statistical applications. For example the Kullback-Leibler divergence between two densities, p and q , is

$$E_p[\ln p - \ln q]$$

which in general does not equal (minus) the divergence between q and p

$$E_q[\ln q - \ln p]$$

as expectations are taken with respect to different densities. Amari demonstrates the strong links between these divergence functions and α -connections.

In Critchley, Marriott and Salmon (1993,1994a), these results are extended to preferred point geometry. A preferred point geodesic distance is not restricted to being symmetric unlike those of a standard metric. Further, it is shown that any divergence function will agree (locally) with the (square root) of a preferred point geodesic distance. More specifically the statistically natural preferred point metric which generates Amari's α -connection structure is shown to have strong links with the Kullback-Leibler divergence.

Intuitively, under suitable regularity conditions, one can view the Kullback-Liebler divergence as a measure of (squared) distance on the (infinite dimensional) space of all

density functions. Consider a parametric family to be a finite dimensional subset of this larger space. The distance between two points in this family could be measured either in the infinite dimensional family or by the path length of a geodesic in the parametric family. If the parametric family is totally flat, discussed above, then a statistically natural preferred point geodesic distance will agree with the (square root) of the Kullback-Leibler divergence. In particular in the linear model, which is totally flat, they both agree with the Fisher geodesic distance used in the previous section. Thus in this case the general ambiguity in possible ways of measuring distance disappears. Again for more details see Critchley, Marriott and Salmon (1994a).

Overall, research into the applications of geodesic statistics is at an early stage but we note two further points. Our recent work shows that the Wald statistic emerges as the leading term in a Taylor expansion of the geodesic squared distance statistic, in all our preferred point metric cases. Some progress has also been made on the difficult practical problem of calculating geodesic distances in parametric models by Minarro and Oller (1990).

7 Conclusions.

This paper has analysed the invariance properties of the Wald statistic using the tools of differential geometry and in particular the geometric concepts of metric tensor and geodesic distance. The Wald statistic has been shown to be a geometrically hybrid quantity and it therefore lacks the desired invariance properties. We demonstrated that this fundamental lack of geometric invariance causes the observed finite sample behaviour of the Wald statistic and geometry has been used to analyse exactly why and where Wald tests based on particular algebraic formulations of the null will fail.

We have proposed a Fisher Geodesic Statistic as an invariant extension of the Wald statistic in the case of testing nonlinear restrictions in the linear model. This statistic satisfies, for the nonlinear case, a finite sample inequality with the Score and Likelihood Ratio statistics that corresponds to the well known inequality between the three classical statistics in the linear case. We have also provided sufficient conditions under which the Wald and FG statistics coincide.

Appendix.

Proof of Proposition 5.1

To choose a real valued function k which completes the coordinate system in the two dimensional case we need to solve the partial differential equation

$$Dk I Dg' = 0.$$

It is convenient to use the method of *characteristic curves* to solve this equation. This reduces the problem to a set of ordinary differential equations. The characteristic curves for this problem will simply be the level sets of the function k . Let $(\theta_1(t), \theta_2(t))$ be any such level sets it is found by solving

$$\left(\frac{d\theta_1}{dt}, \frac{d\theta_2}{dt} \right)' = DgI(\theta)^{-1}.$$

These curves will be well defined in a (tubular) neighbourhood of H_0 . By the uniqueness theorem for the solution of partial differential equations by the method of characteristics the function k will be uniquely defined by its value on the null hypothesis, since each characteristic intersects the null.

Two different solutions k and k^* will both have the same level sets which are the characteristic curves.

Proof of Proposition 5.2

$$f_{11} = [(D_{\theta}g)I(\theta)^{-1}(D_{\theta}g)']^{-1}$$

This expression is clearly invariant to different choices of k -coordinate in all cases.

If $\xi = (g, k')'$ and $\xi^* = (g, k^{*'})'$ are any two Fisher orthogonal parameterisations, by Proposition 5.1 k^* is a function of k alone as they both have the same level sets. Hence, using the chain rule,

$$\left(\frac{dk^*}{d\theta} \right) = \left(\frac{dk^*}{dk} \right) \left(\frac{dk}{d\theta} \right)$$

so that since

$$(f_{22}(\xi^*))^{-1} = \left(\frac{dk^*}{dk} \right) (f_{22}(\xi))^{-1} \left(\frac{dk^*}{dk} \right)'.$$

We see that

$$\frac{f_{22}(g_2, k^*)}{f_{22}(g_1, k^*)} = \frac{f_{22}(g_2, k)}{f_{22}(g_1, k)}$$

is invariant. The expression

$$\frac{f_{22}(g + \delta g, k)}{f_{22}(g, k)} = 1 + \delta g \cdot m(g, k) + O(\delta g^2)$$

where

$$m(g, k) = f_{22}^{-1}(g, k) \left(\frac{\partial}{\partial g} f_{22}|_{(g, k)} \right)$$

shows that $m(g, k)$ is an appropriate measure of the variation in f_{22} . It is, as required, invariant to different choices of the Fisher orthogonal coordinates k . It can be calculated as follows. Once we have solved the differential equation for k we know f_{22} as a function of θ and we also then know the elements of

$$B = \begin{pmatrix} Dg \\ Dk \end{pmatrix} = \begin{pmatrix} \frac{\partial g}{\partial \theta_1} & \frac{\partial g}{\partial \theta_2} \\ \frac{\partial k}{\partial \theta_1} & \frac{\partial k}{\partial \theta_2} \end{pmatrix}$$

Inverting we find

$$\bar{B} = \begin{pmatrix} \frac{\partial \theta_1}{\partial g} & \frac{\partial \theta_1}{\partial k} \\ \frac{\partial \theta_2}{\partial g} & \frac{\partial \theta_2}{\partial k} \end{pmatrix} = \frac{1}{\left\{ \frac{\partial g}{\partial \theta_1} \frac{\partial k}{\partial \theta_2} - \frac{\partial g}{\partial \theta_2} \frac{\partial k}{\partial \theta_1} \right\}} \begin{pmatrix} \frac{\partial k}{\partial \theta_2} & -\frac{\partial g}{\partial \theta_2} \\ -\frac{\partial k}{\partial \theta_1} & \frac{\partial g}{\partial \theta_1} \end{pmatrix}$$

Using the chain rule, we now have all we need to calculate $m(g, k)$ from the equation

$$m(g, k) = \frac{1}{f_{22}(g, k)} \left\{ \frac{\partial f_{22}}{\partial \theta_1} \frac{\partial \theta_1}{\partial g} + \frac{\partial f_{22}}{\partial \theta_2} \frac{\partial \theta_2}{\partial g} \right\}$$

Proof of Proposition 5.3 Consider first $g^A = \beta_1 - \beta_2^{-1}$ so that

$$D_{\beta} g = (1, \beta_2^{-2})'.$$

The equations we have to solve for the characteristic curves $(\beta_1(t), \beta_2(t))'$ are therefore:

$$\frac{d\beta_1}{dt} = a^{-1} \quad \text{and} \quad \frac{d\beta_2}{dt} = b^{-1}\beta_2^{-2}.$$

For any constants c_1 and c_2 , these have solutions:

$$\beta_1(t) = a^{-1}t + c_1 \quad \text{and} \quad \frac{1}{3}\beta_2(t)^3 = b^{-1}t + c_2.$$

Imposing the condition that $\beta(0) \in H_0$ we find that $3c_1^3c_2 = 1$. Eliminating t between these equations, we find that

$$\left(\frac{a}{b}\right)\beta_1(t) - \frac{1}{3}\beta_2(t)^3 = \left(\frac{a}{b}\right)c_1 - c_2.$$

Thus we may take

$$k^A(\beta_1, \beta_2) = d^2\beta_1 - \frac{1}{3}\beta_2^3$$

where $d = \sqrt{\frac{a}{b}}$, since the value of this function of β_1 and β_2 is constant along the curve $(\beta_1(t), \beta_2(t))$ which by definition are the k -constant lines.

Consider next $g^B = \beta_1\beta_2 - 1$ so that $D_{\beta}g = (\beta_2, \beta_1)$ yielding the equations

$$\frac{d\beta_1}{dt} = a^{-1}\beta_2 \quad \text{and} \quad \frac{d\beta_2}{dt} = b^{-1}\beta_1.$$

Writing $\lambda = (\sqrt{ab})^{-1}$, we find that

$$\frac{d(\beta_1 + d^{-1}\beta_2)}{dt} = \lambda(\beta_1 + d^{-1}\beta_2) \quad \text{and} \quad \frac{d(\beta_1 - d^{-1}\beta_2)}{dt} = -\lambda(\beta_1 - d^{-1}\beta_2).$$

Thus, for constants c_1 , and c_2 ,

$$\beta_1 + d^{-1}\beta_2 = 2c_1e^{\lambda t} \quad \text{and} \quad \beta_1 - d^{-1}\beta_2 = 2c_2e^{-\lambda t}$$

That is:

$$\beta_1(t) = (c_1e^{\lambda t} + c_2e^{-\lambda t}) \quad \text{and} \quad \beta_2(t) = d(c_1e^{\lambda t} - c_2e^{-\lambda t})$$

Requiring $\beta(0) \in H_0$, we find $d(c_1^2 - c_2^2) = 1$. Eliminating t between the equations, we find that

$$\beta_1(t)^2 - d^{-2}\beta_2(t)^2 = 4c_1c_2.$$

Here it is convenient to take

$$k^B(\beta_1, \beta_2) = \frac{1}{2}(\beta_1^2 - d^{-2}\beta_2^2).$$

Proof of Proposition 5.4 We have here that

$$f_{11} = \left(a^{-1} \left(\frac{\partial g}{\partial \beta_1} \right)^2 + b^{-1} \left(\frac{\partial g}{\partial \beta_2} \right)^2 \right)^{-1}$$

and

$$f_{22} = \left(a^{-1} \left(\frac{\partial k}{\partial \beta_1} \right)^2 + b^{-1} \left(\frac{\partial k}{\partial \beta_2} \right)^2 \right)^{-1}$$

Consider first $g^A = \beta_1 - \beta_2^{-1}$ for which we can take $k^A = d^2\beta_1 - (\frac{1}{3})\beta_2^3$. Thus

$$B = \begin{pmatrix} 1 & \beta_2^{-2} \\ d^2 & -\beta_2^2 \end{pmatrix}$$

whence

$$\bar{B} = \frac{1}{\{\beta_2^2 + d^2\beta_2^{-2}\}} \begin{pmatrix} \beta_2^2 & \beta_2^{-2} \\ d^2 & -1 \end{pmatrix}$$

Using $f_{22} = b(d^2 + \beta_2^4)^{-1}$, the above expression for \bar{B} , and the general formula for $m(g, k)$ given in Section 5.4, we find:

$$m(g, k) = -\frac{4d^2\beta_2^5}{(d^2 + \beta_2^4)^2}.$$

We obtain the results for g^B in a similar manner.

References

- [1] **Amari, S.** (1985), DIFFERENTIAL-GEOMETRIC METHODS IN STATISTICS, Lecture notes in statistics, No. 28, Berlin, Springer-Verlag.
- [2] **Barndorff-Nielsen O.E.** (1989),PARAMETRIC STATISTICAL MODELS AND LIKELIHOOD, Lecture notes in statistics No. 50, Berlin, Springer-Verlag.
- [3] **Barndorff-Nielsen O.E., D.R.Cox and N.Reid** (1986), "The Role of Differential Geometry in Statistical Theory" *International. Statistical Review*, vol 54, pp 83-96.
- [4] **Barndorff-Nielsen O.E. and D.R. Cox** (1994), INFERENCE AND ASYMPTOTICS Monographs on Statistics and Applied Probability, No.52, Chapman and Hall.
- [5] **Beale E.M.L.** (1960), "Confidence Regions in Nonlinear Estimation" *Jnl. of the Royal Statistical Society, Series B*, vol 22, pps 41-88.
- [6] **Berndt E.R. and N.E.Savin** (1977), "Conflict among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model", *Econometrica*, vol 45, pp 1263-78.
- [7] **Breusch T. and P. Schmidt** (1988), "Alternative Forms of the Wald Test: How Long is a Piece of String ?" *Communications in Statistics, Theory and Methods*, vol 17, pps 2789-2795.
- [8] **Calzolari G. and G. Fiorentini** (1993), "Alternative Estimators of the Standard Tobit Model" *Economic Letters*, 42, pps 5-13 .
- [9] **Chentsov N.N.** (1972), STATISTICAL DECISION RULES AND OPTIMAL INFERENCE (IN RUSSIAN) Nauka, Moscow; translated in to English (1982), AMS, Rhode Island.
- [10] **Cox D.R. and D.V. Hinkley** (1990), THEORETICAL STATISTICS London, Chapman and Hall.
- [11] **Critchley F., P.K. Marriott , and M. Salmon** (1993), " Preferred Point Geometry and Statistical Manifolds", *Annals of Statistics* , vol 21, No.3, pps 1197-1224.
- [12] **Critchley F., P.K. Marriott , and M. Salmon** (1994a), " Preferred Point Geometry and the Local Differential Geometry of the Kullback-Leibler Divergence", forthcoming *Annals of Statistics* .

- [13] Critchley F., P.K. Marriott , and M. Salmon (1994b), " An Elementary Account of Amari's Expected Geometry", *European University Institute working paper*, No. 94/12, submitted for publication.
- [14] Dagenais M.C. and J-M. Dufour (1991), "Invariance, Nonlinear Models and Asymptotic Tests" , *Econometrica*, vol 59, 6, pps 1601-1615.
- [15] Dodson C.T.J. and T. Poston (1977), *TENSOR GEOMETRY*, London, Pitman.
- [16] Evans A.E. and N.E. Savin (1982), "Conflict among the Criteria Revisited; The W, LM and LR Tests" , *Econometrica*, vol 50, pps 737-48.
- [17] Gregory A. and M.R. Veall (1985) "Formulating Wald Tests of Nonlinear Restrictions" , *Econometrica* vol 50, pps 549-563.
- [18] Griffiths W.E., R.C. Hill and P.J.Pope (1987), "Small Sample Properties of Probit Model Estimators", *Jnl. of the American Statistical Association*, vol 82, pps 929-937.
- [19] Gouriéroux C. and A.Monfort (1989), "A General Framework for Testing a Null Hypothesis in a "Mixed" Form" *Econometric Theory* vol 5, No.1, pps 63-82.
- [20] Gouriéroux C., A. Holly and A.Monfort (1982), " Likelihood Ratio Test, Wald Test and Kuhn Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters", *Econometrica*, 50, pps 63-80.
- [21] Hauck and Donner (1977), "Wald's Test as Applied to Hypotheses in Logit Analysis" *Jnl. of the American Statistical Association* , vol 72, pps 851-853.
- [22] Hauck and Donner (1980), "Wald's Test as Applied to Hypotheses in Logit Analysis" *Jnl. of the American Statistical Association*, corrigendum, vol75, .
- [23] Hougaard P. (1982), "Parametrizations of Nonlinear Models " *Journal of the Royal Statistical Society*, Series B, vol 44, No.2, pps 244-252.
- [24] Kass, R.E. (1984), "Canonical Parametrisations and Zero Parameter Effects Curvature" ,*Journal of the Royal Statistical Society* ,Series B, vol 46, pps 86-92.
- [25] Kass, R.E. (1989), "The Geometry of Asymptotic Inference (with discussion)", *Statistical Science* ,4, pps 187-263.
- [26] Kodde D.A. and F.Palm (1986), "Wald Criterion for Jointly Testing Equality and Inequality Restrictions", *Econometrica* , vol 54, pps 1243-1248.

- [27] LeCam L. (1990), "On the Standard Asymptotic Confidence Ellipsoids of Wald" *International Statistical Review*, vol 58, No.2, pps 129–152.
- [28] Lafontaine F. and K.J. White (1986), "Obtaining any Wald Statistic you Want" *Economic Letters*, vol 21, pps 35–40.
- [29] Minarro A. and J.M. Oller (1990), "An Algorithm for Geodesic Coordinates in Riemannian Manifolds", *Indian Jnl. Pure Applied Math.*, vol 21(7), pps 707–715.
- [30] Moolgavkar S.H. and D.J. Venson (1987), "Confidence Regions in Curved Exponential Families: Application to Matched Case Control and Survival Studies with General Relative Risk Function", *The Annals of Statistics*, vol 15, No.1, pps 346–359.
- [31] Murray M.K. and J.W. Rice (1993), DIFFERENTIAL GEOMETRY AND STATISTICS Monographs on Statistics and Applied Probability, No. 48, Chapman and Hall, London.
- [32] Nelson F.D. and N.E. Savin (1986), "On the Invariance Properties of the Wald, LR. and LM. Tests", *Working Paper, Department of Economics, University of Iowa*
- [33] Nelson F.D. and N.E. Savin (1988), "The Non-monotonicity of the Power Function of the Wald Test in Nonlinear Models" *Working Paper, Department of Economics, University of Iowa*
- [34] Phillips P. and J.Y. Park (1988), "On the Formulation of Wald Tests of Nonlinear Restrictions", *Econometrica*, vol 56, No.5, pps 1065–1083.
- [35] Rao C.R., (1945), "Information and Accuracy Attainable in the Estimation of Statistical Parameters." *Bull. Calcutta Math. Soc.*, vol 37, pps 81–91.
- [36] Spivak M. (1981), A COMPREHENSIVE INTRODUCTION TO DIFFERENTIAL GEOMETRY. , Publish or Perish, Berkley, CA.
- [37] Vaeth M (1985), "On the Use of Wald's Test in Exponential Families" *International Statistical Review*, vol 53, No.2, pps 199–214.
- [38] Wolak F.A. (1988), "Duality in Testing Multivariate Inequality Constraints", *Biometrika*, vol 75, pps 611–615.
- [39] Wolak F.A. (1989), "Local and Global Testing of Linear and Nonlinear Inequality Constraints in Nonlinear Econometric Models", *Econometric Theory*, vol 5, No.1, pps 1–35.



EUI WORKING PAPERS

EUI Working Papers are published and distributed by the
European University Institute, Florence

Copies can be obtained free of charge
– depending on the availability of stocks – from:

The Publications Officer
European University Institute
Badia Fiesolana
I-50016 San Domenico di Fiesole (FI)
Italy

Please use order form overleaf



Publications of the European University Institute

Department of Economics Working Paper Series

To Department of Economics **WP**
European University Institute
Badia Fiesolana
I-50016 San Domenico di Fiesole (FI)
Italy

From Name
Address
.....
.....
.....
(Please print)

- ☐ Please enter/confirm my name on EUI Economics Dept. Mailing List
- ☐ Please send me a complete list of EUI Working Papers
- ☐ Please send me a complete list of EUI book publications
- ☐ Please send me the EUI brochure Academic Year 1995/96

Please send me the following EUI **ECO** Working Paper(s):

No, Author
Title:
No, Author
Title:
No, Author
Title:
No, Author
Title:

Date Signature

**Working Papers of the Department of Economics
Published since 1993**

ECO No. 93/1

Carlo GRILLENZONI
Forecasting Unstable and Non-Stationary
Time Series

ECO No. 93/2

Carlo GRILLENZONI
Multilinear Models for Nonlinear Time
Series

ECO No. 93/3

Ronald M. HARSTAD/Louis PHILIPS
Futures Market Contracting When You
Don't Know Who the Optimists Are

ECO No. 93/4

Alan KIRMAN/Louis PHILIPS
Empirical Studies of Product Markets

ECO No. 93/5

Grayham E. MIZON
Empirical Analysis of Time Series:
Illustrations with Simulated Data

ECO No. 93/6

Tilman EHRBECK
Optimally Combining Individual
Forecasts From Panel Data

ECO NO. 93/7

Víctor GÓMEZ/Agustín MARAVALL
Initializing the Kalman Filter with
Incompletely Specified Initial Conditions

ECO No. 93/8

Frederic PALOMINO
Informed Speculation: Small Markets
Against Large Markets

ECO NO. 93/9

Stephen MARTIN
Beyond Prices Versus Quantities

ECO No. 93/10

José María LABEAGA/Angel LÓPEZ
A Flexible Demand System and VAT
Simulations from Spanish Microdata

ECO No. 93/11

Maozu LU/Grayham E. MIZON
The Encompassing Principle and
Specification Tests

ECO No. 93/12

Louis PHILIPS/Peter MØLLGAARD
Oil Stocks as a Squeeze Preventing
Mechanism: Is Self-Regulation Possible?

ECO No. 93/13

Pieter HASEKAMP
Disinflation Policy and Credibility: The
Role of Conventions

ECO No. 93/14

Louis PHILIPS
Price Leadership and Conscious
Parallelism: A Survey

ECO No. 93/15

Agustín MARAVALL
Short-Term Analysis of Macroeconomic
Time Series

ECO No. 93/16

Philip Hans FRANCES/Niels
HALDRUP
The Effects of Additive Outliers on Tests
for Unit Roots and Cointegration

ECO No. 93/17

Fabio CANOVA/Jane MARRINAN
Predicting Excess Returns in Financial
Markets

ECO No. 93/18

Iñigo HERGUERA
Exchange Rate Fluctuations, Market
Structure and the Pass-through
Relationship

ECO No. 93/19

Agustín MARAVALL
Use and Misuse of Unobserved
Components in Economic Forecasting

ECO No. 93/20

Torben HOLVAD/Jens Leth
HOUGAARD
Measuring Technical Input Efficiency for
Similar Production Units:
A Survey of the Non-Parametric
Approach

ECO No. 93/21

Stephen MARTIN/Louis PHILIPS
Product Differentiation, Market Structure
and Exchange Rate Passthrough

ECO No. 93/22

F. CANOVA/M. FINN/A. R. PAGAN
Evaluating a Real Business Cycle Model

ECO No. 93/23

Fabio CANOVA
Statistical Inference in Calibrated Models

ECO No. 93/24

Gilles TEYSSIERE
Matching Processes in the Labour Market
in Marseilles. An Econometric Study

ECO No. 93/25

Fabio CANOVA
Sources and Propagation of International
Business Cycles: Common Shocks or
Transmission?

ECO No. 93/26

Marco BECHT/Carlos RAMÍREZ
Financial Capitalism in Pre-World War I
Germany: The Role of the Universal
Banks in the Financing of German
Mining Companies 1906-1912

ECO No. 93/27

Isabelle MARET
Two Parametric Models of Demand,
Structure of Market Demand from
Heterogeneity

ECO No. 93/28

Stephen MARTIN
Vertical Product Differentiation, Intra-
industry Trade, and Infant Industry
Protection

ECO No. 93/29

J. Humberto LOPEZ
Testing for Unit Roots with the k-th
Autocorrelation Coefficient

ECO No. 93/30

Paola VALBONESI
Modelling Interactions Between State and
Private Sector in a "Previously" Centrally
Planned Economy

ECO No. 93/31

Enrique ALBEROLA ILA/J. Humberto
LOPEZ/Vicente ORTOS RIOS
An Application of the Kalman Filter to
the Spanish Experience in a Target Zone
(1989-92)

ECO No. 93/32

Fabio CANOVA/Morten O. RAVN
International Consumption Risk Sharing

ECO No. 93/33

Morten Overgaard RAVN
International Business Cycles: How
much can Standard Theory Account for?

ECO No. 93/34

Agustín MARAVALL
Unobserved Components in Economic
Time Series *

ECO No. 93/35

Sheila MARNIE/John
MICKLEWRIGHT
"Poverty in Pre-Reform Uzbekistan:
What do Official Data Really Reveal?"

ECO No. 93/36

Torben HOLVAD/Jens Leth
HOUGAARD
Measuring Technical Input Efficiency for
Similar Production Units:
80 Danish Hospitals

ECO No. 93/37

Grayham E. MIZON
A Simple Message for Autocorrelation
Correctors: DON'T

ECO No. 93/38

Barbara BOEHNLEIN
The Impact of Product Differentiation on
Collusive Equilibria and Multimarket
Contact

ECO No. 93/39

H. Peter MØLLGAARD
Bargaining and Efficiency in a
Speculative Forward Market

*out of print

ECO No. 94/1

Robert WALDMANN
Cooperatives With Privately Optimal
Price Indexed Debt Increase Membership
When Demand Increases

ECO No. 94/2

Tilman EHRBECK/Robert
WALDMANN
Can Forecasters' Motives Explain
Rejection of the Rational Expectations
Hypothesis?

ECO No. 94/3

Alessandra PELLONI
Public Policy in a Two Sector Model of
Endogenous Growth *

ECO No. 94/4

David F. HENDRY
On the Interactions of Unit Roots and
Exogeneity

ECO No. 94/5

Bernadette GOVAERTS/David F.
HENDRY/Jean-François RICHARD
Encompassing in Stationary Linear
Dynamic Models

ECO No. 94/6

Luigi ERMINI/Dongkoo CHANG
Testing the Joint Hypothesis of Rational-
ity and Neutrality under Seasonal Coin-
tegration: The Case of Korea

ECO No. 94/7

Gabriele FIORENTINI/Agustín
MARAVALL
Unobserved Components in ARCH
Models: An Application to Seasonal
Adjustment *

ECO No. 94/8

Niels HALDRUP/Mark SALMON
Polynomially Cointegrated Systems and
their Representations: A Synthesis

ECO No. 94/9

Mariusz TAMBORSKI
Currency Option Pricing with Stochastic
Interest Rates and Transaction Costs:
A Theoretical Model

ECO No. 94/10

Mariusz TAMBORSKI
Are Standard Deviations Implied in
Currency Option Prices Good Predictors
of Future Exchange Rate Volatility?

ECO No. 94/11

John MICKLEWRIGHT/Gyula NAGY
How Does the Hungarian Unemploy-
ment Insurance System Really Work?

ECO No. 94/12

Frank CRITCHLEY/Paul
MARRIOTT/Mark SALMON
An Elementary Account of Amari's
Expected Geometry

ECO No. 94/13

Domenico Junior MARCHETTI
Procyclical Productivity, Externalities
and Labor Hoarding: A Reexamination of
Evidence from U.S. Manufacturing

ECO No. 94/14

Giovanni NERO
A Structural Model of Intra-European
Airline Competition

ECO No. 94/15

Stephen MARTIN
Oligopoly Limit Pricing: Strategic
Substitutes, Strategic Complements

ECO No. 94/16

Ed HOPKINS
Learning and Evolution in a
Heterogeneous Population

ECO No. 94/17

Berthold HERRENDORF
Seigniorage, Optimal Taxation, and Time
Consistency: A Review

ECO No. 94/18

Frederic PALOMINO
Noise Trading in Small Markets

ECO No. 94/19

Alexander SCHRADER
Vertical Foreclosure, Tax Spinning and
Oil Taxation in Oligopoly

ECO No. 94/20

Andrzej BANIAK/Louis PHILIPS
La Pléiade and Exchange Rate Pass-
Through

ECO No. 94/21

Mark SALMON
Bounded Rationality and Learning;
Procedural Learning

*out of print

ECO No. 94/22

Isabelle MARET
Heterogeneity and Dynamics of
Temporary Equilibria: Short-Run Versus
Long-Run Stability

ECO No. 94/23

Nikolaos GEORGANTZIS
Short-Run and Long-Run Cournot
Equilibria in Multiproduct Industries

ECO No. 94/24

Alexander SCHRADER
Vertical Mergers and Market Foreclosure:
Comment

ECO No. 94/25

Jeroen HINLOOPEN
Subsidising Cooperative and Non-
Cooperative R&D in Duopoly with
Spillovers

ECO No. 94/26

Debora DI GIOACCHINO
The Evolution of Cooperation:
Robustness to Mistakes and Mutation

ECO No. 94/27

Kristina KOSTIAL
The Role of the Signal-Noise Ratio in
Cointegrated Systems

ECO No. 94/28

Agustín MARAVALL/Víctor GÓMEZ
Program SEATS "Signal Extraction in
ARIMA Time Series" - Instructions for
the User

ECO No. 94/29

Luigi ERMINI
A Discrete-Time Consumption-CAP
Model under Durability of Goods, Habit
Formation and Temporal Aggregation

ECO No. 94/30

Debora DI GIOACCHINO
Learning to Drink Beer by Mistake

ECO No. 94/31

Víctor GÓMEZ/Agustín MARAVALL
Program TRAMO "Time Series
Regression with ARIMA Noise, Missing
Observations, and Outliers" -
Instructions for the User

ECO No. 94/32

Ákos VALENTINYI
How Financial Development and
Inflation may Affect Growth

ECO No. 94/33

Stephen MARTIN
European Community Food Processing
Industries

ECO No. 94/34

Agustín MARAVALL/Christophe
PLANAS
Estimation Error and the Specification of
Unobserved Component Models

ECO No. 94/35

Robbin HERRING
The "Divergent Beliefs" Hypothesis and
the "Contract Zone" in Final Offer
Arbitration

ECO No. 94/36

Robbin HERRING
Hiring Quality Labour

ECO No. 94/37

Angel J. UBIDE
Is there Consumption Risk Sharing in the
EEC?

ECO No. 94/38

Berthold HERRENDORF
Credible Purchases of Credibility
Through Exchange Rate Pegging:
An Optimal Taxation Framework

ECO No. 94/39

Enrique ALBEROLA ILA
How Long Can a Honeymoon Last?
Institutional and Fundamental Beliefs in
the Collapse of a Target Zone

ECO No. 94/40

Robert WALDMANN
Inequality, Economic Growth and the
Debt Crisis

ECO No. 94/41

John MICKLEWRIGHT/
Gyula NAGY
Flows to and from Insured
Unemployment in Hungary

ECO No. 94/42

Barbara BOEHNLEIN

The Soda-ash Market in Europe:
Collusive and Competitive Equilibria
With and Without Foreign Entry

ECO No. 94/43

Hans-Theo NORMANN

Stackelberg Warfare as an Equilibrium
Choice in a Game with Reputation Effects

ECO No. 94/44

Giorgio CALZOLARI/Gabriele
FIORENTINI

Conditional Heteroskedasticity in
Nonlinear Simultaneous Equations

ECO No. 94/45

Frank CRITCHLEY/Paul MARRIOTT/
Mark SALMON

On the Differential Geometry of the Wald
Test with Nonlinear Restrictions

